Bias on estimation in quotient space and correction methods: Applications to statistics on organ shapes

Nina Miolane, Loïc Devilliers, Xavier Pennec

Bias on estimation in quotient space and correction methods

Applications to statistics on organ shapes

Nina Miolane[a,*,**], Loic Devilliers[*] and Xavier Pennec[*]
[*] Université Côte d'Azur and Inria, Epione, 2004 Route des Lucioles, 06902 Valbonne, France
[**] Stanford University, Department of Statistics, Sequoia Hall, 390 Serra Mall, CA 94305 Stanford, United States
[a] Corresponding: nina.miolane@inria.fr

Abstract

Riemannian geometry and the theory of quotient spaces facilitate the analysis of medical imaging algorithms dealing with organ shapes. These algorithms often start with the computation of a *template* organ shape that serves as a reference for normalizing the measurements of each individual data into a common space. The template represents the organ's "prototype" for further analyses. The template is modeled as a parameter of a generative model that is estimated from the observed data, i.e. from noisy images of organs. A usual procedure for template estimation is the computation of the Fréchet mean of the observed data projected in a quotient space. This chapter introduces the geometry of quotient spaces and uses it to show that the usual template estimation procedure is biased. Riemannian geometry allows us to explain the origin of the bias and to design bias correction methods, in order to improve statistical analyses on organ shapes.

9.1. Introduction

The shape of a set of points, the shape of a signal, the shape of a surface, or the shape in an image can be defined as the remainder after we have filtered out the position and the orientation of the object [Ken84]. Studying shapes in medical images has many applications. For example, orthopaedic surgeons analyze *bones' shapes* for surgical pre-planning [DBD+14]. In neuroimaging, studying *brain shapes* as they appear in the MRIs facilitates discoveries on disorders like Alzheimer's disease [LAP15].

What do these applications have in common? Position and orientation of the anatomical structures do not matter for the studies' goal: only *shapes* matter. Mathematically, the study analyzes the statistical distributions of *the equivalence classes of the data* under translations and rotations. This amounts to projecting the data in a quotient space.

The most widely used method for summarizing shape data is the computation of the mean shape. Researchers refer to the mean shape with different terms: mean configuration, mean pattern, template, atlas, etc. We use the term "template" in this chapter. The template is most often computed in practice as a representative of the average of the data equivalence classes. This average is a Fréchet mean in the quotient space and the corresponding procedure is often called the "max-max algorithm" [AAT07].

Author preprint. Chapter 9 of *Riemannian Geometric Statistics in Medical Image Analysis, Pages 343-376.*
X. Pennec, S. Sommer and T. Fletcher editors. Elsevier, 2020. With authorization.
https://doi.org/10.1016/B978-0-12-814725-2.00017-0

We may wonder if the procedure of template estimation is biased. If it is, then inferences in Computational Anatomy on organ shapes will be too. This chapter uses Riemannian geometry and statistics on quotient spaces to show that a bias is indeed introduced under certain conditions. We illustrate the bias in examples from Computational Anatomy and present correction methods.

9.2. Shapes and quotient spaces

9.2.1. Group actions

Consider three-dimensional MRIs of different brains. The concept of the brains' "shapes" has an intuitive meaning: we can make statements such as "these two shapes look different" or "these two shapes look similar". Computational Anatomy seeks to formalize and quantify such statements. Can we give a good *formal definition* of the concept of "shape"? Can we give a good *mathematical representation* of an organ shape? We present a mathematical formalism of shapes by considering the following complementary question: can we give a good representation of what leaves a shape invariant? This section presents the notion of *group actions* as an answer to this question.

Definition 9.1 (Group action of G on a set $\mathcal{M}$)**.** A group (left) action of a group G on a set $\mathcal{M}$ is a map:

$$\begin{aligned} \rho : G \times \mathcal{M} &\to \mathcal{M}, \\ (g, x) &\mapsto g.x \end{aligned}$$

such that for all $x \in \mathcal{M}, e.x = x$, where e is the identity element of G, and for all $(g, h) \in G^2$ and for all $x \in \mathcal{M}, h.(g.x) = (h \circ g).x$. A right group action can be defined similarly.

We use the action of $g \in G$ on an object $x \in \mathcal{M}$ to represent a transformation that leaves the shape of x invariant. In other words, x and $g.x$ have the same shape for any $g \in G$.

Definition 9.2 (Lie group action of G on a differentiable manifold $\mathcal{M}$)**.** If G is a Lie group, $\mathcal{M}$ is a differentiable manifold and ρ is differentiable group action, then ρ is a Lie group action of G on $\mathcal{M}$.

Definitions 9.1 - 9.2 are illustrated in Examples 9.1 - 9.4 from Computer Vision, Signal Processing and Medical Imaging. In these examples, we note that several left and right group actions can be defined on the same space $\mathcal{M}$. The space $\mathcal{M}$ itself can be a set, a finite or infinite dimensional vector space, or a finite or infinite dimen-

sional differentiable manifold. G can be a finite group as well as a finite or infinite dimensional Lie group.

Example 9.1 (k landmarks in $\mathbb{R}^d$)**.** A k-tuple of landmarks in d dimensions is a function ϕ from the space of labels $\{1, ..., k\}$ to the ambient space $\mathbb{R}^d$:

$$\phi : \qquad \{1, ..., k\} \longmapsto \mathbb{R}^d. \tag{9.1}$$

Here, the space of objects is the space of k-tuples, i. e. the space of functions $\mathcal{M} = \mathcal{F}(\{1, .., k\}, \mathbb{R}^d)$ and is finite dimensional. Two groups act on the space of k-tuples $\mathcal{M}$. First, the finite group of permutations $G_1 = \mathfrak{S}$ acts naturally on the space of labels and induces a right action on $\mathcal{M}$. Second, the finite dimensional Lie group $G_2 = \mathrm{SE}(d)$ of rigid body transformations naturally acts on $\mathbb{R}^d$ and induces a left action on $\mathcal{M}$.

Some authors do not consider the right action on the space of labels and work with $\mathcal{M} = \left(\mathbb{R}^d\right)^k$ equipped with a left action of $G = \mathrm{SE}(d)$, for example in Kendall analyses [Ken77] or Procrustean analyses [Goo91, DM98, GD04].

Example 9.2 (k landmarks in S^d)**.** We can also consider k-tuple of landmarks on hyperspheres. Such k-tuple is a function ϕ from the space of labels $\{1, ..., k\}$ to the ambient space S^d:

$$\phi : \qquad \{1, ..., k\} \longmapsto S^d. \tag{9.2}$$

The space of objects is the space of functions $\mathcal{M} = \mathcal{F}(\{1, .., k\}, S^d)$ and is finite dimensional. Two groups act on the space of k-tuples $\mathcal{M}$. The finite group of permutations $G_1 = \mathfrak{S}$ acts naturally on the space of labels and induces a right action on $\mathcal{M}$. The finite dimensional Lie group $G_2 = \mathrm{SO}(d + 1)$ of rotations naturally acts on S^d and induces a left action on $\mathcal{M}$.

Example 9.3 (Continuous 1D signals and 2D/3D images)**.** A continuous 1D-signal is a continuous function ϕ from its support $[a, b]$ to $\mathbb{R}$:

$$\phi : \quad [a, b] \longmapsto \mathbb{R}. \tag{9.3}$$

Here, the space of objects is the space of functions $\mathcal{M} = \mathcal{F}([a, b], \mathbb{R})$ and is infinite dimensional. Several groups act on the space of 1D signals $\mathcal{M}$. The (infinite dimensional) group $\mathrm{Diff}([a, b])$ of diffeomorphisms (also called warpings or reparameterizations) of $[a, b]$ induces a right action on $\mathcal{M}$, see Chapter 4. The (finite dimensional) group of translations and scalings, which naturally acts on $\mathbb{R}$, induces a left action on $\mathcal{M}$ [KSW11]. As another example, a (medical) image is a function ϕ from its domain

D to the grey levels $[0, 256]$:

$$\phi : \quad D \longmapsto [0, 256]. \tag{9.4}$$

Here, the space of objects is the space of functions $\mathcal{M} = \mathcal{F}(D, [0, 256])$. The group of diffeomorphisms $\mathrm{Diff}(D)$ of the domain D induces a right action on $\mathcal{M}$, see again Chapter 4. Subgroups of $\mathrm{Diff}(D)$ are also often considered, like the LDDMM subgroup, or the rotations or affine transformations [LAP15].

Example 9.4 (Discrete 1D signals and 2D/3D images)**.** A 1D discrete signal with periodic boundary conditions is a function ϕ from the set of N points to $\mathbb{R}$:

$$\phi : \quad \mathbb{Z}/N\mathbb{Z} \longmapsto \mathbb{R}. \tag{9.5}$$

This can be generalized to discrete 2D or 3D images, where the values are now intensities:

$$\phi : \quad (\mathbb{Z}/N\mathbb{Z})^d \longmapsto \mathbb{R}. \tag{9.6}$$

When $d = 1$ this is a discrete signal with N points, $d = 2$ a discrete image $N \times N$ pixels and $d = 3$ a discrete 3D image $N \times N \times N$ pixels. The action of translations on the pixels' coordinates is a simplified setting for image registration.

9.2.2. Orbit, isotropy group, quotient space

We consider a set $\mathcal{M}$ equipped with a left action of a group G, denoted $(\mathcal{M}, G, .)$. The action of the group G on $\mathcal{M}$ formalizes the statement "$x \in \mathcal{M}$ and $y \in \mathcal{M}$ have the same shape" through the notion of orbit, that we define now.

Definition 9.3 (Orbit of $x \in \mathcal{M}$)**.** The *orbit of* $x \in \mathcal{M}$ under the action of G, written $[x]$ is defined as:

$$[x] = \{g.x \mid g \in G\}. \tag{9.7}$$

The orbit of $x \in \mathcal{M}$ contains the points of $\mathcal{M}$ reachable with the action of G on x.

Intuitively, the orbit of $x \in \mathcal{M}$ is the set of all objects in $\mathcal{M}$ that have the same shape as x. The relation "having the same shape" defined by $x \sim y : [x] = [y]$ is an equivalence relation whose equivalence classes are the orbits. By properties of the equivalence relations, the orbits defined by this action form a partition of $\mathcal{M}$.

An object $x \in \mathcal{M}$ can be left unchanged by the action of an element $g \in G$. Not only the action of g on x does not change the shape of x, the action does not change the object x at all.

Definition 9.4 (Fixed points of $g \in G$)**.** A point $x \in \mathcal{M}$ is a fixed point of $g \in G$, if:

$$g.x = x. \tag{9.8}$$

Equivalently, we say that g fixes x.

When $x \in \mathcal{M}$ is a fixed point of g, g can be seen as a symmetry of the object's shape. When we consider all the elements g that leave the object x unchanged, i. e. all the symmetries of x, we define the isotropy group of x.

Definition 9.5 (Isotropy group of $x \in \mathcal{M}$)**.** The isotropy group (or stabilizer) of x, written G_x, is defined as the subgroup of transformations that leaves this element fixed:

$$G_x = \{g \in G \mid g.x = x\}. \tag{9.9}$$

Each isotropy group G_x is a subgroup of G.

Remark 9.1 (Special case of a Lie group action)**.** If $(\mathcal{M}, G, .)$ is a Lie group action of G on $\mathcal{M}$, then each orbit $[x]$ is a submanifold of $\mathcal{M}$ and each isotropy group is a Lie subgroup of G.

We are interested in studying the shapes of the objects in $\mathcal{M}$, i. e. we are interested in studying the orbits of objects.

Definition 9.6 (Quotient space)**.** The set of orbits is called the *quotient space of* $\mathcal{M}$ *by the action of* G:

$$\mathcal{Q} = \mathcal{M}/G = \{[x] \mid x \in \mathcal{M}\}. \tag{9.10}$$

Intuitively, the quotient space $\mathcal{Q}$ is the space of the shapes of the objects described in $\mathcal{M}$. $\mathcal{Q}$ is sometimes called the shape space.

Remark 9.2 (Terminology)**.** The spaces $\mathcal{M}$ and $\mathcal{Q}$ have different names in the literature. $\mathcal{M}$ can be called the top space, the ambient space, the object space, etc. $\mathcal{Q}$ can be called the bottom space, the quotient space, the shape space, etc.

Definitions 9.3, 9.5 and 9.6 are illustrated in Examples 9.5-9.6 and in Figure 9.1.

Example 9.5 (2 landmarks in the plane $\mathbb{R}^2$)**.** Consider two landmarks in the plane $\mathbb{R}^2$, one red and one black as in Figure 9.1 (left). The landmarks are initially parameterized each with 2 coordinates. We then consider that one landmark is fixed at the origin on $\mathbb{R}^2$. Thus the system is now parameterized by the 2 coordinates of the second landmark

only, e.g. in polar coordinates (r, θ). We consider the action of the Lie group $SO(2)$ on the second landmark. This action does not change the shape of the system of the two landmarks. The shape of the 2 landmarks is the distance between them which is simply r in our notations: the shape is an element of the quotient space $\mathbb{R}^2/SO(2)$.

Figure 9.1 (left) shows the action of g on the landmark x by a blue arrow and the orbit $[x]$ by a blue dotted circle. The shape space is the space $Q = \mathbb{R}_+$ of all possible distances between the two landmarks. Every $x \neq (0, 0)$ has isotropy group the identity and $(0, 0)$ has isotropy group the whole group of 2D rotations.

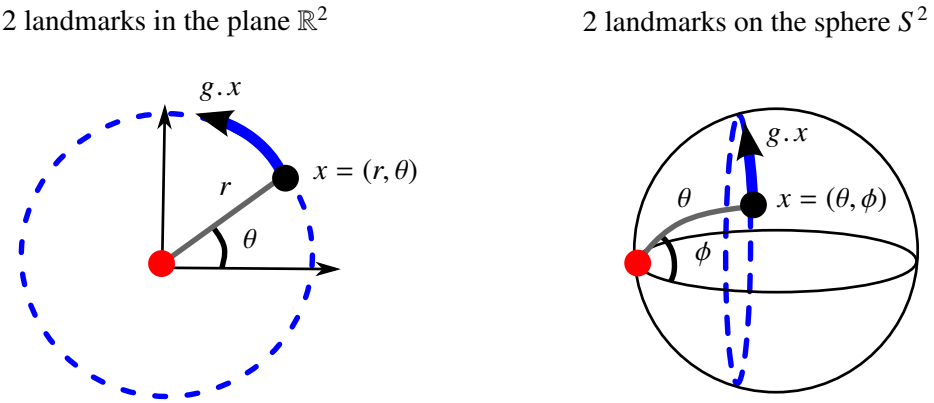

Figure 9.1 2 landmarks, one in red and one in black, in the plane $\mathbb{R}^2$ (left) and on the sphere $\mathbb{S}^2$ (right) described in Examples 9.5 - 9.6. The blue arrow shows the action of $g \in \mathrm{SO}(2)$ on the landmark x and the blue dotted circle represents the orbit of x.

Example 9.6 (2 landmarks on the sphere $\mathbb{S}^2$)**.** Consider two landmarks on the sphere $\mathbb{S}^2$, one in red and one in black as in Figure 9.1 (right). One of the landmarks is fixed at a pole of $\mathbb{S}^2$. The system is now parameterized by the 2 coordinates of the second landmark only, i.e. $x = (\theta, \phi)$, where θ is the latitude and ϕ the longitude. We consider the action of the Lie group $SO(2)$ on the second landmark. This action does not change the shape of the system of the two landmarks. The shape of the two landmarks is the angle between them, which is simply θ in our notations: the shape is an element of the quotient space $\mathbb{S}^2/SO(2)$.

Figure 9.1 (right) shows the action of g on the landmark x by a blue arrow and the orbit in $\mathcal{M}$ by a blue dotted circle. The shape space is $Q = [0, \pi]$, the space of all possible angles between the two landmarks.

9.2.3. Proper and effective group actions

There are different types of group actions which have different properties. A group action can be: proper, effective (or faithful), free, locally free, transitive, regular (or simply transitive or sharply transitive), etc. From now on, we consider only proper and effective actions and define them below.

Definition 9.7 (Proper group action)**.** A group action of G on $\mathcal{M}$ is proper if the map:

$$\begin{aligned} \varphi \colon G \times \mathcal{M} &\to \mathcal{M} \times \mathcal{M} \\ (g, x) &\to (g.x, x) \end{aligned}$$

is proper, i.e. if the pre-image of any compact set is compact.

If G is compact and if the map φ is continuous, then the action of G on $\mathcal{M}$ is proper. A proper action enables to separate points in the quotient space, i. e. to separate shapes in the shape space, as shown in the Proposition below.

Proposition 9.1. *If the action of G on $\mathcal{M}$ is proper, then every orbit is a closed subset of $\mathcal{M}$ and the quotient space $\mathcal{Q}$ is Haussdorff.*

Definition 9.8 (Effective group action)**.** A group action of G on $\mathcal{M}$ is effective if:

$$\cap_x G_x = \{e\}, \tag{9.11}$$

i.e. if the only group element that leaves all points fixed is the identity element. An effective group action is also called faithful.

Lastly, we consider the situation where there are no fixed points for any $g \in G$, which is the definition of a free action given below.

Definition 9.9 (Free group action)**.** A group action of G on $\mathcal{M}$ is free if all isotropy groups are trivial, i.e. equal to $\{e\}$.

The actions described in Examples 9.5 - 9.6 are proper and effective, but not free. For example, there are points fixed by all elements of the group: their isotropy groups are not trivial, they are equal to the whole group.

9.2.4. Principal and singular orbits

The isotropy group of an object x controls the "size" of its orbit, i. e. the amount of objects that have the same shape as x:

Theorem 9.1 (Orbit-stabilizer theorem)**.** *For $x \in \mathcal{M}$, consider the map from G to $\mathcal{M}$ given by: $g \mapsto g.x$. Its image is the orbit $[x]$. Besides, the following map defined by:*

$$G/G_x \to [x]$$
$$g \circ G_x \mapsto g.x$$

is well defined and is a bijection.

If $x, y \in \mathcal{M}$ are in the same orbit $[x] = [y]$, then their isotropy groups G_x and G_y are conjugate groups in G, i.e. there exists an element $g \in G$ such that: $G_x = g \circ G_y \circ g^{-1}$.

Definition 9.10 (Orbit type)**.** The *orbit type* of the orbit $[x]$ is defined as the conjugacy class (H) of the isotropy group G_x.

The "smaller" is the isotropy group, the "larger" is the orbit and its type: if a shape has less symmetries, then there exist more configurations of objects that have the same shape. The definition below formalizes the notion of ordering that can be defined between isotropy groups and corresponding orbit types.

Definition 9.11 (Partial ordering on isotropy groups and orbit types)**.** Let H and K be isotropy subgroups of G, with isotropy types denoted by (H) and (K). We write $(H) \leq (K)$ if and only if H is conjugate to a subgroup of K.

This defines a partial ordering on the set of isotropy groups and orbit types. The "largest" orbit type exists and is unique under the principal orbit theorem [AKLM03].

Theorem 9.2 (Principal orbit theorem)**.** *Consider $(\mathcal{M}, G, .)$ a proper Lie group action of G on a connected differential manifold $\mathcal{M}$. Then, the smallest orbit type is unique and the stratum generated by principal orbits is open and dense in $\mathcal{M}$. We denote it $\check{\mathcal{M}}$.*

Under the assumptions of the above theorem, we can give the following definitions.

Definition 9.12 (Principal and singular orbits)**.** The principal orbit type is the orbit type with largest isotropy group. We call principal orbits the orbits with principal orbit type. We call singular orbits any other orbit.

We call principal shapes the shapes that correspond to the principal orbits: they are "non-degenerated" shapes. Similarly, singular shapes correspond to singular orbits:

they are shapes with symmetries, i. e. degenerated shapes. The density of $\check{\mathcal{M}}$ means that there are objects with principal shapes almost everywhere.

The following proposition gives additional properties of sets of orbits with same orbit type and of $\check{\mathcal{M}}$ in particular [AKLM03].

Proposition 9.2. *Consider an isotropy type and the set $\mathcal{N}$ of orbits with this isotropy type. Then $\mathcal{N}$ and $\mathcal{N}/G$ are smooth manifolds. Furthermore, the inclusion $\mathcal{N}/G \to \mathcal{M}/G$ is smooth and the submersion $\mathcal{N} \to \mathcal{N}/G$ is smooth.*

Under our assumptions, the principal shape space $\check{\mathcal{Q}} = \check{\mathcal{M}}/G$ is a smooth manifold and $\check{\pi} : \check{\mathcal{M}} \to \check{\mathcal{Q}}$ is a smooth submersion. $\check{\mathcal{Q}}$, the space of non-degenerated shapes, is the manifold part of $\mathcal{Q}$.

How do the singular shapes, or singularities of $\mathcal{Q}$, or singular orbits of $\mathcal{M}$, enter the picture? Consider $(\mathcal{M}, G, .)$ a Lie group action of G on a differentiable manifold $\mathcal{M}$. The orbits - principal or singular - can be gathered into a stratification of $\mathcal{M}$, as formalized below.

Definition 9.13 (Stratification and orbit type stratification)**.** A *stratification of $\mathcal{M}$* is a locally finite partition of $\mathcal{M}$ by embedded submanifolds called *strata*, required to fit together in a certain way, called the *frontier condition*. The connected components of the *orbit types* form a stratification of $\mathcal{M}$, called the *orbit-type stratification of $\mathcal{M}$*.

The orbit-type stratification induces a stratification of the quotient space.

Proposition 9.3. *The quotient space $\mathcal{Q}$ is a stratified space, where the strata are the connected components of the orbit types.*

Example 9.5 continued (2 landmarks in the plane $\mathbb{R}^2$)**.** The principal stratum of $\mathcal{M} = \mathbb{R}^2$ is $\check{\mathcal{M}} = \mathbb{R}^2 \setminus (0,0)$ which is dense in $\mathbb{R}^2$. Likewise, principal shapes form an open and dense subset $\check{\mathcal{Q}} = \mathbb{R}_+^*$ which is dense in $\mathcal{Q} = \mathbb{R}_+$.

Example 9.6 continued (2 landmarks on the sphere $\mathbb{S}^2$)**.** The principal stratum in $\mathcal{M}$ is $\check{\mathcal{M}} = \mathbb{S}^2 \setminus \{(0,0),(\pi,0)\}$ which is dense in $\mathbb{S}^2$. The point $(0,0)$ denotes one pole of $\mathbb{S}^2$ and $(\pi,0)$ its opposite pole in $\mathbb{S}^2$. In Q, the principal stratum is $\check{\mathcal{Q}} =]0,\pi[$ which is dense in $\mathcal{Q} = [0,\pi]$.

9.2.5. Metric structure

We now formalize and quantify the statement: “these shapes are similar”. So far, we did not introduce any notion of distance or dissimilarity between objects or shapes.

We address this by adding a Riemannian structure on the previous framework.

We consider a vector space or a differentiable manifold $(\mathcal{M}, G, .)$ equipped with a group action. Let $d_{\mathcal{M}}$ be a distance defined on $\mathcal{M}$, that is a way to quantify the (dis)similarity between shapes. We study below the "compatibility" of this distance with the group action.

Definition 9.14 (G-invariant distance and isometric action)**.** The distance $d_{\mathcal{M}}$ on $\mathcal{M}$ is invariant under the action of the group G on $\mathcal{M}$, and is called a G-invariant distance, if:

$$\forall g \in G, \quad \forall x, y \in \mathcal{M} \quad d_{\mathcal{M}}(x, y) = d_{\mathcal{M}}(g.x, g.y). \tag{9.12}$$

The distance between two elements in $\mathcal{M}$ is conserved after an action by the same group element g. Equivalently, the action is said to be isometric with respect to $d_{\mathcal{M}}$ because it conserves the distances.

The distance $d_{\mathcal{M}}$ may come from an inner product or a Riemannian metric on $\mathcal{M}$. Intuitively, a G-invariant distance means that the (dis)similarity between two objects does not change if we transform both objects in the same way.

Example 9.5 continued (2 landmarks in the plane $\mathbb{R}^2$)**.** Consider the Euclidean plane $(\mathbb{R}^2, \langle\ ,\ \rangle)$ where $\langle\ ,\ \rangle$ is the canonical inner product. The action of $SO(2)$ is isometric with respect to the distance induced by $\langle\ ,\ \rangle$.

Example 9.6 continued (2 landmarks on the sphere $\mathbb{S}^2$)**.** Consider the sphere $(\mathbb{S}^2, \langle\ ,\ \rangle)$ where $\langle\ ,\ \rangle$ is the Riemannian metric on $\mathbb{S}^2$ induced by the canonical inner product in $\mathbb{R}^3$. The action of $SO(2)$ is isometric with respect to the Riemannian distance induced by $\langle\ ,\ \rangle$.

Example 9.4 continued (Discrete 1D signals and 2D/3D images)**.** Consider the space of discrete 1D signals or the space of discrete 2D/3D image with the distance being the sum of the square differences in intensities. The action of translation on coordinates is isometric with respect to this distance.

Remark 9.3 (Special case of a Hilbert space $(\mathcal{M}, \langle\ ,\ \rangle)$)**.** We consider the distance in the Hilbert space $\mathcal{M}$, given by the norm: $d_{\mathcal{M}}(a, b) = \|a - b\|$. We say that G acts isometrically and linearly on $\mathcal{M}$, if $x \mapsto g.x$ is a linear map which leaves the norm unchanged.

Remark 9.4 (Special case of a Riemannian manifold $(\mathcal{M}, \langle\ ,\ \rangle)$)**.** The Riemannian

metric on $\mathcal{M}$ is G-invariant if the differential of its action ρ:

$$d\rho_g : T_x\mathcal{M} \to T_{g.x}\mathcal{M} \tag{9.13}$$

leaves the metric $\langle\ ,\ \rangle_x$ at x invariant. A G-invariant Riemannian metric induces a G-invariant Riemannian distance.

The distance in $\mathcal{M}$ represents a measure of (dis)similarity between objects. We examine how to turn it into a distance in $\mathcal{Q}$: a measure of (dis)similarity between shapes. A first step in this direction is the concept of registration of objects, also called alignment, which can be formulated using group actions:

Definition 9.15 (Optimal positioning)**.** We say that the point $g.x_1$ is in optimal position to x_2 if:

$$d_{\mathcal{M}}(g.x_1, x_2) = \inf_{g\in G} d_{\mathcal{M}}(g.x_1, x_2). \tag{9.14}$$

Equivalently, we say that the objects x_1 and x_2 are registered (Medical imaging terminology, e.g. [MV96]) or aligned (Signal processing terminology, e.g. [KSW11]).

Then, a distance on $\mathcal{M}$ induces a distance on the quotient space $\mathcal{Q}$:

Definition 9.16 (Distance on the quotient space $\mathcal{Q}$)**.** Consider $d_{\mathcal{M}}$ be a distance on $\mathcal{M}$ and $(\mathcal{M}, G, .)$ an isometric action with respect to $d_{\mathcal{M}}$. Then:

$$d_{\mathcal{Q}}([x],[y]) = \inf_{g\in G} d_{\mathcal{M}}(g.x, y) \tag{9.15}$$

is a distance on $\mathcal{Q}$.

Example 9.1 continued (k landmarks in $\mathbb{R}^d$)**.** Consider the case of k-tuples in $\mathbb{R}^d$ with the action of the rigid body motions $\mathrm{SE}(d)$ on $(\mathbb{R}^d)^k$ of Example 9.1. The Euclidean metric on $\mathbb{R}^d$ induces a metric on $(\mathbb{R}^d)^k$ which is invariant for the action of $\mathrm{SE}(d)$. This induces a distance on the quotient space, which is computed in practice by first registering the k-tuples and then using the distance on $(\mathbb{R}^d)^k$: the Procrustean distance [Sma96, DM98]. The Riemannian structure of the quotient space has been studied in by Le and Kendall [LK93].

In practice, the distance in $\mathcal{Q}$, i.e. the distance between two shapes, is computed by first registering the objects, and then using the distance in the ambient space $\mathcal{M}$. The registration is a crucial step as there is a priori no closed formed expression to compute the distance between two points in $\mathcal{Q}$, even if there is one in $\mathcal{M}$. One case where it may be possible is when an isometric section exists.

Definition 9.17 (Global and local isometric sections)**.** Take $\pi : \mathcal{M} \mapsto Q$ the canonical projection into the quotient space. Let $\mathcal{U}$ be an open subset of Q. A map $s : \mathcal{U} \mapsto \mathcal{M}$ is a *local section* if $\pi \circ s = \mathrm{Id}$. Moreover, we say that s is *isometric* if

$$\forall o, o' \in \mathcal{U}, \quad d_Q(o, o') = d_{\mathcal{M}}\left(s(o), s(o')\right). \tag{9.16}$$

Then the image $\mathcal{S} = s(\mathcal{U})$ of the open subset $\mathcal{U}$ of the quotient by the section s is a subset of $\mathcal{M}$ with the following property: $\forall x, y \in \mathcal{S}, d_{\mathcal{M}}(x, y) = d_Q([x], [y])$. Moreover we say that the section is global if $\mathcal{U} = Q$.

A global section gives a subset of $\mathcal{M}$ containing a point of each orbit such that all points in $\mathcal{S}$ are registered. A global isometric section rarely exists.

Example 9.7. There exists a global isometric section for the Examples 9.5 and 9.6 of the two landmarks in $\mathbb{R}^2$ and $\mathbb{S}^2$: we can compute a closed form expression for the distance in the respective quotient spaces $\mathbb{R}_+$ and $[0, \pi]$. However, we can show that there is no global isometric section for Example 9.4 of discrete 1D signals and 2D/3D discrete images.

9.3. Template estimation

Differential Geometry on quotient spaces, introduced in the previous sections, gives a mathematical framework to analyze algorithms in Medical Imaging and Computational Anatomy. We consider the algorithms of template estimation. Intuitively, a template of a given population is a prototype of this population. In Medical Imaging for example, a template of a data base of brain images is a brain image representing a reference anatomy. The template shows the prototype shape of the brain population under study.

Various methods exist to compute a template from a given database, see [DM98] for templates of landmarks or [EJCB12] for brain templates. A first practice was to select one object from the database as the template. If the selected object's shape is far from the population mean shape, the template is necessarily biased towards this specific data point. Thus, the template fails at being a prototype of the population. Therefore, researchers have developed other algorithms to compute the template. We investigate here the computation of the template as a Fréchet mean in the quotient space.

9.3.1. Generative model

The template represents a prototype of the data. We formally define it here as a parameter of a generative model. We have a sample $\{X_i\}_{i=1}^n$ of size n, which can be sets of landmarks, curves, images, etc. Each element of the sample is interpreted

as a noisy observation of the template up to a group action that does not change the template's shape, for example the action of a repositioning or a reparameterization [AAT07, DATP17b, BG10, BC11, KSW11]. We present below common generative models, first for data in a Hilbert space then for data in a Riemannian manifold.

Observations in a Hilbert space

Let $(\mathcal{M}, , \langle\ ,\ \rangle, G, .)$ be a (potentially infinite dimensional) Hilbert space with a group action. The generative model of the data is defined as:

$$X = g.y_0 + \epsilon, \tag{9.17}$$

where $y_0 \in \mathcal{M}$ is the template, g is a random transformation in G and ϵ is a random variable that represents a standardized noise in $\mathcal{M}$ with null mean $\mathbb{E}(\epsilon) = 0$ and finite variance $\mathbb{E}(\|\epsilon\|^2) < \infty$. We assume that g and ϵ are independent random variables.

Example 9.5 continued (2 landmarks in the plane $\mathbb{R}^2$)**.** The generative model of the landmarks in the 2D plane writes in Cartesian coordinates:

$$(x, y) = R_\theta.(0, r_0) + \epsilon = r_0(\cos\theta, \sin\theta) + \epsilon.$$

where $y_0 = (0, r_0)$ is the template, the 2D rotation matrix R_θ of rotation angle θ is a random variable in $SO(2)$ and ϵ is the noise in $\mathbb{R}^2$.

Observations in a finite dimensional Riemannian manifold

Let $(\mathcal{M}, \langle\ ,\ \rangle, G, .)$ be a finite dimensional Riemannian manifold with a group action. $\mathrm{Exp}_p(u)$ denotes the Riemannian exponential of u at point p. The generative model of the data is defined as:

$$X = \mathrm{Exp}_{g.y_0}(\epsilon), \tag{9.18}$$

where $y_0 \in \mathcal{M}$ is the template, $g \in G$ is a random transformation and ϵ is a random variable in the tangent space $T_{g.y_0}\mathcal{M}$ which represents the noise. We assume that g and ϵ are independent random variables. We assume the noise to be isotropic Gaussian of standard deviation σ in each coordinate on the tangent space $T_{g.y_0}\mathcal{M}$.

Example 9.6 continued (2 landmarks on the sphere $\mathbb{S}^2$)**.** The generative model writes in latitude and longitude coordinates:

$$(\lambda, \phi) = R_\theta.(y_0, 0) + \epsilon = (y_0, \theta) + \epsilon, \tag{9.19}$$

where $y_0 = (y_0, 0)$ is the template, the random 2D rotation matrix R_θ is a random variable in SO(2) and ϵ is the noise.

An alternative generative model where the noise is added before the group action

can also be considered for data belonging respectively to a Hilbert and to a Riemannian manifold:

$$X = g.(y_0 + \epsilon) \quad \text{resp.} \quad X = g.\mathrm{Exp}_{y_0}(\epsilon). \tag{9.20}$$

A model without noise $X = g.y_0$ can be found in [KSW11].

9.3.2. An iterative estimation procedure

The template $y_0 \in \mathcal{M}$ is defined as a parameter of the generative model. We now describe the usual procedure of template estimation [KSW11, AAT07, DATP17b, BG10, BC11, JDJG04], which is given in Algorithm 9.1. Let $(\mathcal{M}, \langle\ ,\ \rangle, G, .)$ be a Hilbert space or a finite dimensional Riemannian manifold equipped with a group action of G. Data are generated from the generative model described in the previous subsection. The template estimate is initialized with one of the observed data $\hat{y} = X_1$. Then, the procedure iterates two steps (i) and (ii) until convergence of the template estimation.

Algorithm 9.1 Template estimation

Input: Observed data $\{X_i\}_{i=1}^n$, convergence threshold δ

Initialization: $k = 0$ and $\hat{y}_0^{(0)} = X_1$

Repeat:

- (i) Registration: $\forall i \in \{1, ..., n\}, \hat{g}_i = \underset{g \in G}{\mathrm{argmin}}\ d_{\mathcal{M}}(\hat{y}_0^{(k)}, g.X_i)$,
- (ii) Fréchet mean computation: $\hat{y}_0^{(k+1)} = \underset{y \in \mathcal{M}}{\mathrm{argmin}}\ \sum_{i=1}^n d_{\mathcal{M}}(y, \hat{g}_i.X_i)^2$
- $k \leftarrow k + 1$

until convergence: $d_{\mathcal{M}}(\hat{y}_0^{(k)}, \hat{y}_0^{(k-1)}) < \delta$

Output: $\hat{y}_0^{(k)}$.

Step (i) is the registration of each object X_i to the current template. We assume that each minimizer $\hat{g}_i$ exists and is reached by (i). In practice, this is the case when G is compact. Step (ii) is the computation of the Fréchet mean of the registered data $\hat{g}_i.X_i$. We assume that the minimizer $\hat{y}_0^{(k+1)}$ exists and is reached in (ii). In practice, this is the case for low levels of noise σ in the generative model, as the registered data $\hat{g}_i.X_i$ will end-up being concentrated on a small neighbourhood of $\mathcal{M}$ [EM91]. The procedure described in Algorithm 9.1 is sometimes called the "max-max procedure".

Example 9.5 continued (2 landmarks in the plane $\mathbb{R}^2$)**.** Let x_1, x_2, x_3 be three objects in $\mathbb{R}^2$ in Figure 9.2 (left). Step (i) filters out the position or parameterization component, i. e. the coordinate on the orbit. The objects x_1, x_2, x_3 are projected in the shape space Q using the blue arrows. Then, the Fréchet mean is computed in Step (ii).

Example 9.6 continued (2 landmarks on the sphere $\mathbb{S}^2$)**.** Similarly, Figure 9.2 shows Steps (i) and (ii) of the template estimation procedure for a sample of size 3, where each element is a set of 2 landmarks on the sphere $\mathbb{S}^2$.

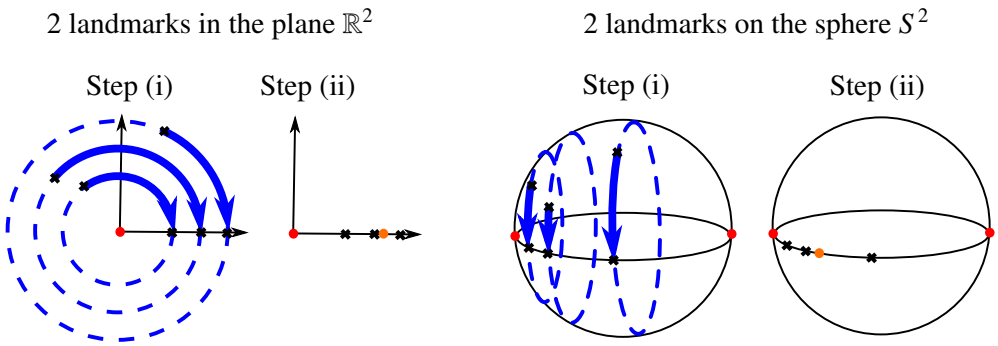

Figure 9.2 Steps (i)-(ii) of template estimation. The 3 black plus signs in $\mathbb{R}^2$ (left) or $\mathbb{S}^2$ (right) represent the 3 data. The 3 dotted blue curves are their orbits. In Step (i), the data are registered. The 3 black crosses in $\mathbb{R}_+$ (positive x-axis) (left) or $[0, \pi]$ (right) represent the registered data. In Step (ii), the Fréchet mean of the registered data is computed and shown in orange.

9.3.3. Convergence to the Fréchet mean in the quotient space

We consider the convergence of the iterative procedure of template estimation, described in Algorithm 9.1, for $k \to +\infty$ where k is the number of iterations. The procedure decreases at each step the following cost, which is bounded below by zero:

$$\text{Cost}(g_1, ..., g_n, y) = \sum_{i=1}^{n} d^2_{\mathcal{M}}(y, g.X_i). \tag{9.21}$$

Under the assumptions that both steps (i) and (ii) reached their minimizers, we are guaranteed convergence to a local minimum.

Proposition 9.4. *We assume that the procedure converges to the global minimum for $k \to +\infty$. Then, the estimate of the template is:*

$$\hat{y}_0 = \underset{y \in \mathcal{M}}{argmin} \sum_{i=1}^{n} \min_{g \in G} d^2_{\mathcal{M}}(y, g.X_i). \tag{9.22}$$

where we recognize the sample Fréchet mean in the quotient space, as in Section 4.6.2 of Chapter 4. Moreover, if the group is finite, the algorithm converges in a finite

number of steps [DATP17a].

We recall that the template y_0 is defined as an element of $\mathcal{M}$. Proposition 9.4 shows that the iterative procedure converges to $\hat{y}_0$ which is an element of the quotient space Q. Recovering y_0 is indeed an ill-posed problem. We can only estimate the equivalence class of y_0. In other words, we estimate the shape of the template.

We want to compare $\hat{y}_0$ to the parameter y_0 it was designed to estimate: to this aim, we want y_0 and $\hat{y}_0$ to be in the same space. We assume that there exists a local isometric section s around y_0 and compare $s(\hat{y}_0)$, a representative object of the shape $\hat{y}_0$, to the template y_0 in $\mathcal{M}$. In the following, we write $\hat{y}_0$ or $s(\hat{y}_0)$ indifferently, as well as $\pi(y_0)$ or y_0 indifferently, since the local geometry of $\check{Q}$ is equivalent to the local geometry of its local isometric section in $\check{\mathcal{M}}$.

9.3.4. Other convergence(s)

We can investigate two other types of convergence in the estimation procedure. First, we consider the convergence in the sample size $n \to +\infty$. The sample Fréchet mean on a manifold converges to the population Fréchet mean set when sample size goes to infinity [Zie77]. Thus, we assume that we have an infinite sample and we consider the population Fréchet mean as the estimator of the template:

$$\hat{y}_0 = \underset{y \in \mathcal{M}}{\operatorname{argmin}} \int \min_{g \in G} d^2_{\mathcal{M}}(y, g.X)\, d\mathcal{M}(X). \tag{9.23}$$

Second, we consider the convergence as the noise level goes to 0: $\sigma \to 0$. When there is no noise, the population Fréchet mean in Q gives the template shape of the generative model: the estimator gives the parameter it was designed to estimate. There is no bias in the estimation in this case. The next sections investigate what happens when the noise is non-zero, i.e. in the context of a real experimental setting.

9.3.5. Bias of the procedure

Consider the template y_0 and a representative of its estimate $\widehat{y_0}$. We want to know if the procedure described above has a bias, when is noise is non-zero: $\sigma \neq 0$. The following definition generalizes to Riemannian manifolds the usual definition of a bias $\text{Bias}(\hat{y_0}, y_0) = \mathbb{E}\left[\hat{y_0} - y_0\right]$ in a linear space.

Definition 9.18 (Bias)**.** Consider $(\mathcal{M}, \langle\ ,\ \rangle)$ a Riemannian manifold and Log its Riemannian logarithm. Take $y_0 \in \mathcal{M}$ and $\hat{y_0}^{(n)} \in \mathcal{M}$ an estimator of y_0, computed from a sample of size n. We assume that $\hat{y_0}^{(n)}$ is within the injectivity radius of the Riemannian exponential at y_0 so that the Riemannian logarithm at y_0 is well defined. The bias

of the estimator $\hat{y_0}^{(n)}$ with respect to the (manifold-valued) parameter y_0 is defined as:

$$\mathrm{Bias}(\hat{y_0}^{(n)}, y_0) = \mathbb{E}\left[\mathrm{Log}_{y_0}(\hat{y_0}^{(n)})\right]. \tag{9.24}$$

where the expectation is taken over the repeated draws of samples of size n. The asymptotic bias is defined as the bias of the estimator $\hat{y_0}^{(\infty)}$ with respect to the parameter y_0.

From now on, we assume that $\hat{y_0}^{(n)}$ converges in probability to a single value $\hat{y_0}^{(\infty)} = \hat{y}_0$, when the sample size goes to infinity. If this value is exactly the original template value y_0, then the estimate is said to be consistent.

Definition 9.19 (Weak consistency)**.** A estimator is weakly consistent if it converges in probability to the parameter it is designed to estimate, as the number of data points goes to infinity.

In the following, we use the term "consistency" to refer to the "weak consistency" above. In the general case, the estimation $\hat{y}_0$ differs from the original template value y_0. The asymptotic bias

$$\mathrm{Bias}(\hat{y_0}, y_0) = \mathrm{Log}_{y_0}(\hat{y_0}) \tag{9.25}$$

measures how much one would have to shoot from y_0, along a Riemannian geodesic, to get the estimated parameter $\widehat{y_0}$.

Remark 9.5 (Variance and asymptotic variance)**.** In statistics, an estimator is evaluated based on its bias and variance. Since we assumed that the estimator $\hat{y_0}^{(\infty)}$ of the template converges in probability to a value $\hat{y}_0$, the asymptotic variance of the estimator is vanishing:

$$\mathrm{Var}(\hat{y_0}^{(\infty)}) = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{dist}^2(\hat{y_0}^{(n)}, \hat{y}_0)\right] = 0. \tag{9.26}$$

Thus, we focus on the asymptotic bias.

9.4. Asymptotic bias of template estimation

We first consider $\mathbb{R}^2$ with the canonical inner product as an example of a finite dimensional Hilbert space and $\mathbb{S}^2$ with the metric inherited from $\mathbb{R}^3$ as an example of a finite dimensional Riemannian manifold. We show the asymptotic bias of the template estimation for these special cases, before extending to general cases.

9.4.1. Intuition on examples

The special cases of 2 landmarks on the plane $\mathbb{R}^2$ and on the sphere $\mathbb{S}^2$, introduced in Examples 9.5-9.6, are useful to show the origin of the asymptotic bias of $\hat{y}_0$. We consider a generative model with a Gaussian isotropic noise of standard deviation σ on each coordinate. As long as $\sigma \neq 0$, there is *a bias that comes from the curvature of the template's orbit.* Figure 9.3 shows the template's orbit and the level set σ of the Gaussian noise for both examples.

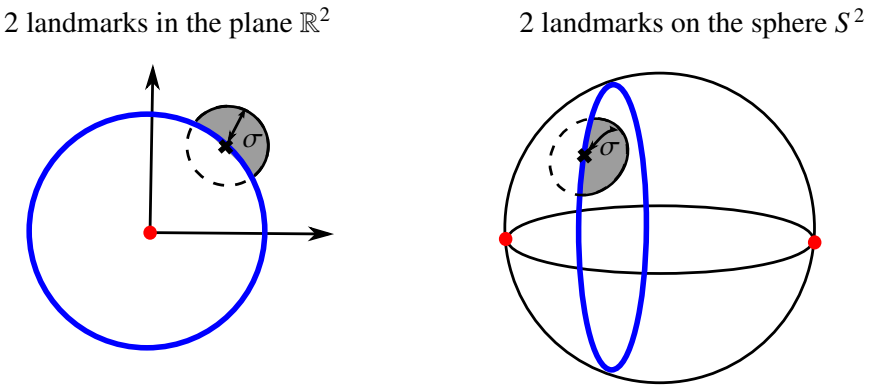

Figure 9.3 Geometric origin of the bias. The extrinsic curvature of the template's orbit creates the asymptotic bias. The blue curves represent the respective templates' orbits. The balls of radius σ represent a level set of the Gaussian noise distribution. The grey-colored areas represent the noise distribution that generates data "outside" the orbit of y_0, i. e. on the orbit side that is the furthest away from the closest singularity.

Example 9.5 continued (2 landmarks in the plane $\mathbb{R}^2$)**.** Figure 9.3 (left) shows that the probability of generating an observation X_i outside of the template's shape orbit is larger than the probability of generating it inside: the grey area in the black circle is larger than the white area. When the data are registered, and projected in the shape space, there will be more registered data that are greater than the template. Their expected value, which is the template estimator, will therefore be greater than the template: there is a bias of the estimator with respect to the parameter it is designed to estimate.

Figure 9.4 (left) shows the bias of $\hat{y}_0$ with respect to y_0, as a function of σ. Increasing the noise level σ takes the estimate $\hat{y}_0$ away from y_0. The estimate is driven away from 0: it goes to ∞ when $\sigma \to \infty$. One can show numerically that the bias varies as σ^2 around $\sigma = 0$.

Example 9.6 continued (2 landmarks on the sphere $\mathbb{S}^2$)**.** Figure 9.3 (right) shows that if the template's shape orbit is defined by a constant $\theta < \pi/2$, the probability of generating an observation X_i "outside" of it, i.e. with $\theta_i > \theta$, is larger than the probability of generating it "inside". When the data are registered, and projected in the shape space, there will be more registered data that are greater than the template θ and again, their expected value will also be greater than the template. Conversely, if the template is $\theta > \pi/2$, the phenomenon is inverted: there will be more registered data that are smaller than the template. The average of these registered data will also be smaller than the template. Finally, if the template's shape orbit is the great circle defined by $\theta = \pi/2$, then the probability of generating an observation X_i on the left is the same as the probability of generating an observation X_i on the right. In this case, the registered data will be well-balanced around the template $\theta = \pi/2$ and their expected value will be $\pi/2$: there is no asymptotic bias in this particular case.

Figure 9.4 (right) shows the bias of $\hat{y}_0$ with respect to y_0, as a function of σ. Increasing the noise level σ takes the estimate $\hat{y}_0$ away from y_0. It is repulsed from 0 and π: it goes to $\pi/2$ when $\sigma \to \pi$, as the probability distribution becomes uniform on the sphere in this limit. One can show numerically that the bias varies as σ^2 around $\sigma = 0$.

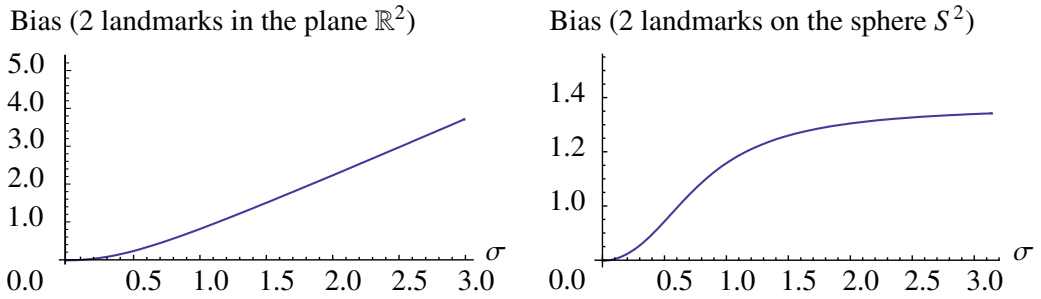

Figure 9.4 Asymptotic bias on the template estimate $\hat{y}_0$ with respect to the noise level σ for $r = 1$ (left) and $\theta = 1$ (right). The bias is quadratic near $\sigma = 0$. Increasing σ takes the estimate $\hat{y}_0$ away from the singularities of Q.

9.4.2. Bias on quotient of finite dimensional Riemannian manifold

This section uses geometry of quotient spaces to prove the origin of the bias observed in the special cases above. Consider a finite dimensional manifold with an isometric proper effective action of a finite dimensional Lie group $(\mathcal{M}, \langle\ ,\ \rangle, G, .)$. The data X_i's are generated in $\mathcal{M}$ through the model 9.18: $X = \mathrm{Exp}_{g.y_0}(\epsilon)$, where g and ϵ are independent random variables. We assume that the template y_0 is in the principal

stratum of $\mathcal{M}$, i.e. $y_0 \in \check{\mathcal{M}}$ and that there exists a local isometric section[1] around y_0. We assume that the noise ϵ is a random variable with isotropic Gaussian distribution of standard deviation σ in each coordinate of the tangent space $T_{g.y_0}\mathcal{M}$.

We show that there is an asymptotic bias on the template shape estimation under these assumptions. We proceed in two steps. The template shape estimate is the Fréchet mean of the observations projected in the quotient space. Thus, we first compute the distribution of the observations in the quotient space Q, see Theorem 9.3. Then, we compute the estimator $\hat{y}_0$ as the expectation of the distribution in Q and compare it with the parameter y_0 to get the bias, see Theorem 9.4.

Theorem 9.3. *[Induced probability density on Q [MHP17]] Under the above assumptions, the probability density function f on the $[X_i]$'s, $i = 1...n$, in the asymptotic regime on an infinite number of observations $n \to +\infty$, has the following Taylor expansion around the noise level $\sigma = 0$:*

$$f(z) = \frac{1}{C_Q(\sigma)} \exp\left(-\frac{d^2_{\mathcal{M}}(y_0, z)}{2\sigma^2}\right)\Big(F_0(z) + \sigma^2 F_2(z) + O(\sigma^4) + \Xi(\sigma)\Big)$$

where z denotes a point in $\check{\mathcal{M}}$ belonging to a local isometric section around y_0. In the above expression, $C_Q(\sigma)$ is the integration constant of the Gaussian. Then, for a fixed z, Ξ is a function of σ that decreases exponentially for $\sigma \to 0$. F_0 and F_2 are functions of z involving the local geometry of $\check{\mathcal{M}}$ around each $z \in \check{\mathcal{M}}$.

The expression of f is obtained by locally integrating the probability distribution function of the observations in $\mathcal{M}$ along the orbits, see [MHP17] for the proof. Thus, f is the probability density function on $\check{Q}$, i.e. the probability density of the object shapes associated to the generative model. Moreover, we have: $d^2_{\mathcal{M}}(y_0, z) = d^2_Q([y_0], [z])$ for $d_{\mathcal{M}}(z, y_0) \leq r$.

Consider now that $\hat{y}_0$ is computed with the template estimation procedure presented in this section. This means that $\hat{y}_0$ is the expectation of the distribution f above. The following theorem computes its Riemannian logarithm from the parameter y_0, i.e. the asymptotic bias of the template shape estimation.

Theorem 9.4 (Asymptotic bias [MHP17])**.** *Under the above assumptions, in the regime of an infinite number of observations $n \to +\infty$, the asymptotic bias of the template shape estimator $\hat{y}_0$, with respect to the parameter y_0, has the following Taylor expan-*

[1]This technical assumption due to the structure of the proof currently limits the application range of the bias analysis of this section. We conjecture that is can be notably softened.

sion around the noise level $\sigma = 0$:

$$Bias(\hat{y}_0, y_0) = -\frac{\sigma^2}{2} H(y_0) + O(\sigma^3) + \epsilon(\sigma), \tag{9.27}$$

where H is the mean curvature vector of the template's orbit which represents the extrinsic curvature of the orbit in $\check{\mathcal{M}}$, and ϵ is a function of σ that decreases exponentially for $\sigma \to 0$.

The proof of this result can be found in [MHP17]. This expression of the bias generalizes the quadratic behavior observed in the examples on Figure 9.4. The asymptotic bias has a geometric origin: it comes from the extrinsic curvature of the template's orbits, see Figure 9.3.

The results are valid when $\mathcal{M}$ is a finite dimensional manifold and G a finite dimensional Lie group. Some interesting examples belong to the framework of infinite dimensional manifolds with infinite dimensional Lie groups. This is the case for the LDDMM framework on images [JKSM06]. Therefore it is important to extend these results to the infinite dimensional case. The next subsection presents results for (infinite dimensional) Hilbert spaces.

9.4.3. Bias on quotient of (in)finite dimensional Hilbert space

Consider a Hilbert space with a linear and isometric group action $(\mathcal{M}, \langle\ ,\ \rangle, G, .)$. We assume that the data are generated with the model: $X = g.(y_0 + \epsilon)$ where $y_0 \in \mathcal{M}$ is the template and ϵ is an independent noise. We define the random variable $Y = y_0 + \epsilon$.

When we are in a Hilbert space of infinite dimension $\mathcal{M}$, it does not make sense to take an isotropic Gaussian variable defined as a Gaussian noise on each coordinate with a parameter σ. Therefore we have to proceed differently to control the noise level: we consider a standardized noise η with null mean and unit variance ($\mathbb{E}(\eta) = 0$ and $\mathbb{E}(\|\eta\|^2) = 1$) and scale it with the standard deviation τ: $\epsilon = \tau\eta$. Then the observable variable is assumed to be:

$$X = g.y_0 + \epsilon = g.y_0 + \tau\eta.$$

When $\mathcal{M}$ is a linear space of finite dimension k, an isotropic Gaussian of parameter σ on each coordinate has the standard deviation $\tau = \sqrt{k}\sigma$. Therefore, σ and τ are two equivalent measures of the noise level which is added as far as we remain in finite dimension.

Theorem 9.5 (Sufficient condition leading to a bias)**.** *Let G be a group acting isometrically and linearly on a Hilbert space $\mathcal{M}$ and $X = g.(y_0 + \epsilon)$ be a random variable in $\mathcal{M}$ with $E(\|X\|^2) < +\infty$. We pose $Y = y_0 + \epsilon$ with $E(Y) = y_0 \neq 0$. If:*

$$\mathbb{P}\left(d_Q([y_0], [Y]) < \|y_0 - Y\|\right) > 0, \tag{9.28}$$

then $[y_0]$ *is not a Fréchet mean of* $[X]$ *in* $Q = \mathcal{M}/G$*. The max-max template estimation procedure is inconsistent.*

Notice that here the linearity assumption is crucial. For instance, this does not hold with translations. In practice it is easy to fulfill the condition given by (9.28). For instance a Gaussian noise fulfills this condition as soon as the template is not a fixed point under the group action. We have also the following proposition:

Proposition 9.5. *Let G be a group acting isometrically and linearly on a Hilbert space* $\mathcal{M}$*, and let Y a random variable in* $\mathcal{M}$ *with* $\mathbb{E}(\|Y\|^2) < +\infty$*. Assume* $Y = y_0 + \epsilon$*, where* $y_0 \neq 0$ *and* $\mathbb{E}(\epsilon) = 0$*. We suppose that* $[y_0]$ *is a sub-manifold of* $\mathcal{M}$ *with tangent space* $T_{y_0}[y_0]$ *at* y_0*. If:*

$$\mathbb{P}(\epsilon \notin T_{y_0}[y_0]^{\perp}) > 0, \tag{9.29}$$

then (9.28) *is fulfilled, and the template estimation is inconsistent.*

For a consistent estimation, we need the support of the noise to be included into a proper linear space of $\mathcal{M}$. This is a severe restriction.

Asymptotic bias for a very large noise

When τ goes to infinity, we have the following a behaviour of the bias:

Theorem 9.6. *Let G be a group acting isometrically and linearly on a Hilbert space* $\mathcal{M}$*. We assume that the support of the noise* ϵ *is not included in the set of fixed points under the group action. Let* $X = g.y_0 + \tau\epsilon$ *be the observable variable. If the Fréchet mean of* $[X]$ *exists, then the asymptotic has the following asymptotic behavior when the noise level* τ *goes to infinity:*

$$Bias(\hat{y}_0, y_0) = \sigma K + o(\tau) \quad as \quad \tau \to +\infty, \tag{9.30}$$

where $K = \sup_{\|v\|=1} \mathbb{E}\left(\sup_{g\in G} \langle v, g.\epsilon\rangle\right) \in (0, 1]$ *is a constant which depends only on the standardized noise and of the group action. In particular, K does not depend on the template.*

9.5. Applications to statistics on organ shapes

The procedure of template shape estimation has an asymptotic bias: even with an infinite number of observations, the estimator does not converge to the parameter it was designed to estimate. We show in this section how the results of Section 9.4 impact the analysis of landmarks and brain shapes.

9.5.1. Shapes defined by landmarks

We first consider landmarks' shapes and we mention the theory introduced by Kendall in the 1980's [Ken77]. Kendall considered shapes of k labelled landmarks in $\mathbb{R}^d$ and assumed that data were observed directly in the quotient space. Thus, the framework of Section 9.4 does not apply in this context: such studies do not consider that the data are observed with noise in the space of landmarks $(\mathbb{R}^d)^k$ and then projected in the shape space. The generative model is different and the question of the bias that we investigate in this chapter is not raised in this context.

However, Section 9.4 applies directly to Procrustean analyses of landmarks shapes [Goo91, DM98, GD04] which consider observations in the space of landmarks. Data are projected in the shape space by an "alignment" or "registration" step. In this literature, the bias has been observed in different settings depending on the assumptions defining the shape, as described below.

Different types of Procrustean analyses

Procrustean analyses that relate to the framework of Section 9.4 are the ones called "Generalized" Procrustean analyses. The original orthogonal Procruste problem was to find the rotation minimizing the residuals between two matrices [Sch66] (other solutions for that problem were proposed with different names [Mos39, Joh66]). "Ordinary" Procrustean analysis is the problem of registering a set of matrices to a reference matrix chosen in advance. "Generalized" Procrustean analysis refers to the problem of group-registration of a set of matrices using the max-max algorithm, i.e. Algorithm 9.1, which is also called the alternation framework in this literature [Gow75, TB77].

Then, the terminology depends on the group of transformations G used to register the matrices. If only rotation matrices are used, this is orthogonal Procrustean analysis. If rotations, translations and scalings (i.e. similarities) are used, this is called "Extended Orthogonal" Procrustean analysis. The setting that relates to the isometric action's framework of Section 9.4 is the use of rotations and translations only.

A distinction has been made in this literature between "shape" and "form" (or size-and-shape) to help determining which group of transformation G is used for registration. "Form" relates to the quotient of the object by rigid body transformations only. "Shape" denotes the quotient of the object by similarities. Kendall shape spaces refer to "shape": the scalings are quotiented by constraining the size of the landmarks' set to be 1.

Note that the group of transformations used to register the objects can be distinct from the group used to generated them in the generative model. If there is no scaling in the generative model, and we estimate the mean form, then we get the partial Procrustes estimate of form. If there is some scaling in the generative model and we

estimate the mean form, we get the full Procrustes estimate of shape [KM97]. The partial estimate of shape can also be considered.

Mean form and mean shape for 2D/3D landmarks

Existing results on bias for shapes of landmarks in 2D and 3D first confirm the findings of Section 9.4, but also give intuition to prove bias results in settings not covered in Section 9.4.

We first consider the Gaussian noise assumption on the landmarks in 2D and 3D. The partial Procrustes estimate of the mean "form" through Generalized Procrustean analysis was known to be inconsistent, as shown in [Lel93] with an reducto ad absurdum proof. Section 9.4 provides a geometric interpretation of this fact, while extending the results for landmarks in higher dimensions. The full Procrustes estimate of the mean "shape" through Generalized Procrustean analysis was known to be consistent for shapes of landmarks in 2D and 3D [Lel93, Le98]. This result goes beyond the settting of Section 9.4 and shows that there can be no bias for a non-isometric action in this specific case.

Now we relax the Gaussian noise assumption on the landmarks. We investigate if there is bias in other noise settings. For landmarks in 2D and under isotropic noise model (non necessarily Gaussian), the full Procrustes estimate of shape is proven to be consistent [KM97]. Under isotropic Gaussian noise model, the partial Procrustes estimator of form is inconsistent, but its shape is correct. Bias thus appears only in the size of the estimate and is shown to be of order σ^2. Then, for non-isotropic errors, the full Procrustes estimate does not need to be consistent and can be arbitrarily inconsistent for high level of noise. This analysis gives an excellent line of work in order to extend the results of Section 9.4 to other noise settings.

9.5.2. Brain images

We turn in this section to shapes of brains as shown in medical images and especially in MRIs. We show how Section 9.4 can be applied to gather intuition about bias of brain template estimation.

Brain template

Computing a brain template is often the first step in neuroimaging studies. The template is often called an atlas in this literature. In general, the brain template is used as a standardized 3D coordinate frame where the subject brains can be compared. The subjects are then characterized by their *spatial diffeomorphic deformations from the template*, see Chapter 4. These deformations may then serve in a statistical analysis of the subject shapes [AHF+98] where the normal and pathological variations of the subjects with respect to the template are quantified. In other words, the template serves as prior knowledge of the brain anatomy [MCAG93]. It should be representative of the

population under study, thus avoiding bias in subsequent statistical analyzes.

Such an "unbiased" template is often constructed by performing an iterative averaging of intensities and deformations [GMT00, JDJG04, HHLAP16] in the spirit of Algorithm 9.1: the brain template can be interpreted as an instance of a Fréchet mean in a quotient space, see Section 4.6.2 of Chapter 4. We can gain insight about its statistical properties using the intuition developed in this chapter and quantify its asymptotic bias, locally on the brain template image.

Spatial bias on the brain template image

We produce maps showing the local asymptotic bias with a color code superimposed on the original tridimensional brain template image, as shown in Figure 9.5. We call these maps the *asymptotic bias maps*. A green color indicates a low asymptotic bias for a given brain region and a red color indicates a high asymptotic bias on another brain region. These maps are computed by leveraging the geometric understanding of this chapter to quantify the bias on the brain image through the following heuristic.

Section 9.4 shows that the asymptotic bias of the template estimate $\hat{y}_0$ depends on the noise level σ at the scale of a distance d to the singularity of the space. The variables d and σ depend on the topology of the brain template image's level sets and can be expressed using the Morse-Smale complex of the image [MHP18]. The indicator quantifying the local asymptotic bias is expressed in a logarithmic scale, as:

$$\mathrm{SNR}_{\mathrm{dB}} = 10 \log_{10} \left[\left(\frac{d}{\sigma} \right)^2 \right]. \tag{9.31}$$

The scale is thus in dB, as the decibel is the logarithmic unit that expresses the ratio of two values of a physical quantity. This unit emphasizes that the quantification of the asymptotic bias depends on a signal-noise ratio (SNR). Indeed, we can consider that the signal is d, which is the template's distance to a singularity and the "noise" is σ, the intersubject variability after registration. The lower is the SNR on a given brain region, the larger is the local asymptotic bias on the brain template on that brain region.

We compute several local asymptotic bias maps for the same brain template, see (c)-(d)-(e) on Figure 9.5. The difference between the maps is a threshold used to compute the asymptotic bias, which is increased from (c) to (e) [MHP18]. The threshold corresponds to the intensity scale at which we look at the images. It controls the spatial scale of the brain regions. Increasing the threshold makes more regions appear, these regions are smaller as well as more biased: they become colored in orange-red on Figure 9.5 (c)-(e).

The interpretation with respect to neuroimaging is the following. Each map of Figure 9.5 (c)-(d)-(e) represents the asymptotic bias of the brain template we would obtain if we were constraining the image's level sets' topology complexity at a given

intensity scale. The local asymptotic bias maps show smaller regions, in orange-red, where the estimated template's brain structures are small with respect to the subjects' variability in the database. In these orange-red regions, it is not reasonable to have a sharply defined template, because the estimated "anatomical" structures may have appeared by chance, by registration of noise between the different subjects. In other words, the maps reveal brain regions where the assumption of a unique anatomy in the subject population may break down.

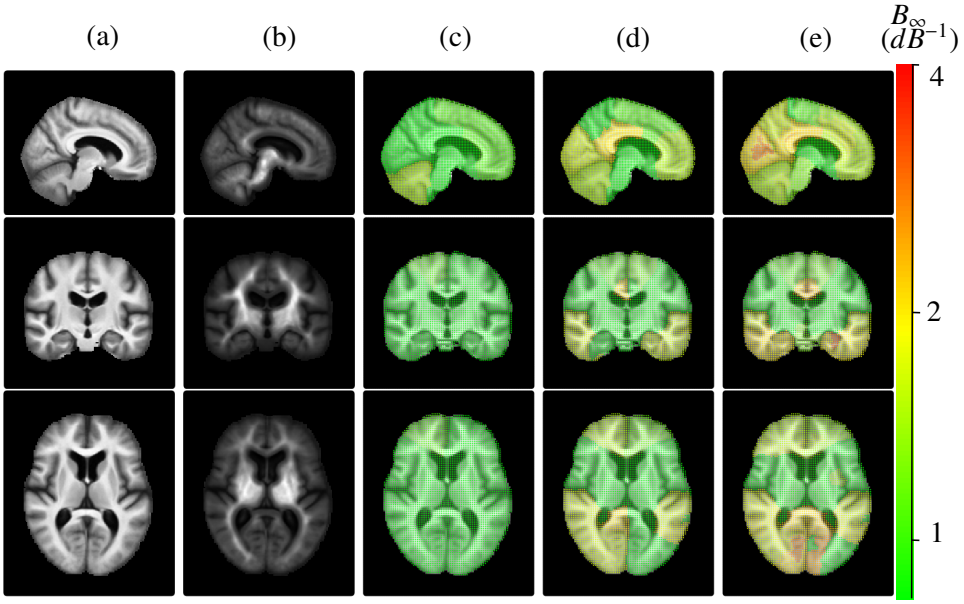

Figure 9.5 Local asymptotic bias maps of the brain template's computed from the OASIS database [MWP+07]. (a) Template. (b) Template whitened by the intersubject variability. (c) Local asymptotic bias maps for a threshold= 1.3, (d) for threshold = 2, (e) for threshold = 4 (dimensionless) [MHP18].

9.6. Bias correction methods

The bias of the template shape estimation impacts analyses of organ shapes in Medical Imaging and Computational Anatomy. We now investigate how to correct for the bias. A first correction would be to change the estimate of the template, moving away from the Fréchet mean in the quotient space. This is for instance what is provided by the maximum likelihood template estimation techniques or Bayesian mixed effect models [ADK15]. This is beyond the scope of this chapter. Here, we present methods to quantify the bias in the Fréchet mean estimator and to correct it when it is sufficiently

important.

9.6.1. Riemannian bootstrap

Procedures to correct the asymptotic bias on the template's estimator are described in [MHP17]. They rely on the bootstrap principle, more precisely on a parametric bootstrap, which is a general Monte Carlo based resampling method that enables one to estimate the sampling distributions of estimators [Efr79]. We focus here on one of the methods, called the iterative bootstrap (Algorithm 9.2), and we refer to [MHP17] for the other method, called the nested bootstrap.

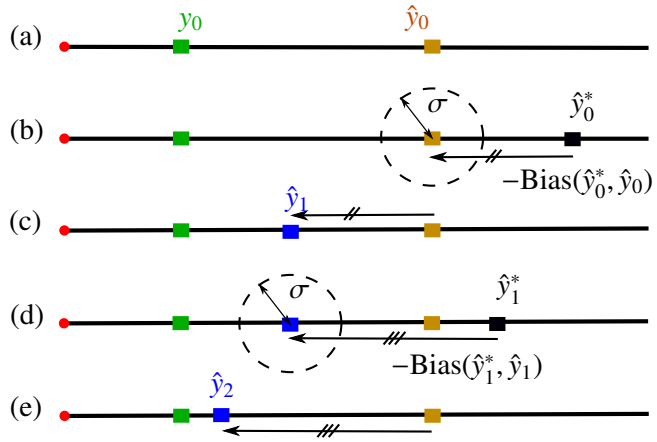

Figure 9.6 Algorithm 9.2 Iterative bootstrap procedure on the example of 2 landmarks in the plane for $n \to +\infty$. (a) Initialization, (b) Generate bootstrap sample from $\hat{y}_0$ and compute the corresponding estimate $\hat{y}_0{}^*$, compute the bias $\hat{y}_0 - \hat{y}_0^*$, (c) Correct $\hat{y}_0$ with the bias to get $\hat{y}_1$, (d) Generate bootstrap sample from $\hat{y}_1$ and iterate as in (b), (e) Get $\hat{y}_2$ etc.

Algorithm 9.2 starts with the usual (biased) template's estimate $\hat{y}_0$, see Figure 9.6 (a), and iteratively improves it. At each iteration, we correct $\hat{y}_0$ with a better approximation of the bias. First, we generate bootstrap data by using $\hat{y}_0$ as the template shape of the generative model. We perform the template's estimation procedure with the Fréchet mean in the quotient space. This gives an estimate $\hat{y}_0^*$ of $\hat{y}_0$. The bias of $\hat{y}_0^*$ with respect to $\hat{y}_0$ is $\text{Bias}(\hat{y}_0^*, \hat{y}_0)$. It gives an estimation $\widehat{\text{Bias}(\hat{y}_0, y_0)}$ of the bias $\text{Bias}(\hat{y}_0, y_0)$, see Figure 9.6 (b). We correct $\hat{y}_0$ by this approximation of the bias. This gives a new estimate $\hat{y}_1$, see Figure 9.6 (c). We recall that the bias $\text{Bias}(\hat{y}_0, y_0)$ depends on y_0, see Theorem 9.4. $\hat{y}_1$ is closer to the template y_0 than $\hat{y}_0$. Thus, the next iteration gives a better estimation $\text{Bias}(\hat{y}_1^*, \hat{y}_1)$ of $\text{Bias}(\hat{y}_0, y_0)$. We correct the initial $\hat{y}_0$ with this better estimation of the bias, etc. The procedure is written formally for a general manifold $\mathcal{M}$:

In Algorithm 9.2, $\Pi_{x_1}^{x_2}$ denotes the parallel transport from $T_{x_1}\mathcal{M}$ to $T_{x_2}\mathcal{M}$ along a geodesic. For linear spaces like $\mathbb{R}^2$ in the plane example, we have $\mathrm{Log}_{x_1} x_2 = \overrightarrow{x_1 x_2}$, $\mathrm{Exp}_{x_1}(u) = x_1 + u$, and the parallel transport is the identity $\Pi_{x_1}^{x_2}(u) = u$. For general manifolds, the parallel transport $\Pi_{x_1}^{x_2}(u)$ can be approximately computed using the Schild's ladder [EPS12] or more accurately using the pole ladder [MP13] (see Chapter 5 of this book). In constant curvature spaces like spheres, the pole ladder is even exact in one step [Pen18].

Algorithm 9.2 is a fixed-point iteration $y_{k+1} = F(y_k)$ where:

$$F(x) = \mathrm{Exp}_{\hat{y}_0}(-\Pi_x^{\hat{y}_0}(\mathrm{Bias})) \qquad \text{and:} \qquad \mathrm{Bias} = \mathrm{Log}_x \hat{x}. \tag{9.32}$$

In a linear setting we have simply $F(x) = \hat{y}_0 - \overrightarrow{x\hat{x}}$. One can show that F is a contraction and that the template shape y_0 is the unique fixed point of F, see [MHP17] for details. Thus the procedure converges to y_0 in the case of an infinite number of observations $n \to +\infty$.

The bootstrap methods work for any type of data, but may be computationally expensive as the estimation procedure is performed at each step of the bootstrap. These methods also depend on how confident we are in the generative model. As such, we can consider alternatives to correct the bias.

9.6.2. Brain images: topologically constrained template

We present a correction method that can specifically be applied to the case of brain images presented in Section 9.5.

Algorithm 9.2 Riemannian iterative bootstrap

Input: Objects $\{X_i\}_{i=1}^n$, noise variance σ^2, convergence threshold δ
Initialization:
$\hat{y}_0 = \text{Fréchet}(\{[X_i]\}_{i=1}^n)$
$k \leftarrow 0$
Repeat:

- Generate bootstrap sample $\{X_i^{(k)^*}\}_{i=1}^n$ from $\mathcal{N}_M(y_k, \sigma^2)$
- $\widehat{y_k} = \text{Fréchet}(\{[X^{(k)^*}]_i\}_{i=1}^n)$ $\mathrm{Bias}_k = \mathrm{Log}_{y_k} \widehat{y_k}$
- $\hat{y}_k = \mathrm{Exp}_{\hat{y}_0}\left(-\Pi_{\hat{y}_k}^{\hat{y}_0}(\mathrm{Bias}_k)\right)$
- $k \leftarrow k + 1$

until convergence: $\|\mathrm{Log}_{\hat{y}_{k+1}} \hat{y}_k\| < \delta$
Output: $\hat{y}_k$

Constrain the topology to control the bias

In Section 9.5, each map of Figure 9.5 (c)-(d)-(e) represents the asymptotic bias of the brain template we would obtain if we were constraining the level sets of the image to have a specific topology of increasing complexity. On the one hand, a complex topology implies an important asymptotic bias on the template, which may not represent faithfully the brain anatomy shared by the subjects in the database. On the other hand, a topology that is too simple has no chance of representing a brain anatomy at all. If we want to look at small brain structures, we have to allow for some precision in the topology.

Which topology shall we choose in this trade-off of asymptotic unbiasedness versus image sharpness? If the local intensity of the computed template is below the noise, there is no hope to compute a consistent template. If the noise is of the same order of magnitude as the signal, the template may estimate the noise instead of the signal. Thus we choose an threshold that expresses the limit situation where signal (intensity on the brain image) and noise are of the same order of magnitude. In practice, this threshold is between -1 and 0 dB [MHP18].

Applying topological denoising on the brain template's

We have decided on the complexity of the brain template topology. This topology can be enforced by applying a topological denoising step in the template estimation procedure [JWS12, GJR$^+$14]. By doing so, we force the asymptotic bias to be below a threshold. Such a control of the brain template's bias enables us to build a template in which the topological denoising step has blurred the image where the sharply defined brain template does not make sense as a representative of the shared brain anatomy [MHP18].

We could be interested in a brain template that would be sharp *and* unbiased. In this case, we could consider dropping the assumption of a unique brain anatomy expressed by a unique template y_0 in the generative model. We could consider multiple templates, expressing the generative model as a mixture model. Further work is needed to investigate the construction of a stratified template, which would add a new stratification every time a region's asymptotic bias crosses the threshold $B_\infty \sim 1dB$.

9.7. Conclusion

This chapter has introduced tools of Riemannian geometry to study the properties of template shape estimation in Medical Imaging. The study of consistency and asymptotic bias that was presented summarizes the results of several works. In particular, Table 9.1 summarizes the contents of the Theorems 9.3-9.6 taken from these works, with their main differences.

In both analyses, the fixed points of the group actions play a fundamental role. The

	Work in [DATP17b, DATP17a]	Work in [MHP17]
Top space $\mathcal{M}$	Hilbert space	Finite dimensional manifold
Noise ϵ	Sufficient condition	Gaussian noise
Quantification (noise level)	Taylor expansion for $\sigma \to +\infty$ and bounds	Taylor expansion for $\sigma \to 0$ with orbit's geometry

Table 9.1 Summary of the mathematical results presented in this chapter.

fixed points are the "worse" singularities of the space, in the sense that their isotropy group is the whole acting group. They are responsible for the curvature of the orbits as shown in the Riemannian manifold case, and they play a role in the hypothesis on the noise's support in the Hilbert case.

Non isometric actions are often encountered in the literature. However, obtaining results in this setting is much more difficult since the distances are now changed by the action of the group. As a consequence, one can not define *a priori* a quotient distance anymore. However, for Hilbert spaces, it is still possible to define the classical loss function:

$$F(m) = \mathbb{E}\left(\inf_{g \in G} \|g.m - X\|^2\right).$$

In this setting, one can show that the template used to generate the variable X does not minimize this function F, at least when the noise level σ is large enough [DATP17a]. Future works could probably improve this result.

We have also presented specific examples of template computations in which bias appears, as well as methods to correct it. Computations of templates have been used in the Medical Imaging literature for at least 15 years. In neuroimaging in particular, computing the template is often the first step of a study and understanding the associated bias is key. Still, a biased template can be seen as an indication that the assumption of a unique template, e.g. a unique brain anatomy within the population, should be relaxed. Further work will investigate the estimation of a mixture of templates or stratified templates and how this may allow to reduce the bias.

Acknowledgment

This work was partially supported by the National Science Foundation, grant NSF DMS RTG 1501767.

Bibliography

[AAT07] Stéphanie Allassonnière, Yali Amit, and Alain Trouvé. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society*, 69(1):3–29, 2007.

[ADK15] Stéphanie Allassonnière, Stanley Durrleman, and Estelle Kuhn. Bayesian Mixed Effect Atlas Estimation with a Diffeomorphic Deformation Model. *SIAM Journal on Imaging Sciences*, 8(3):1367–1395, January 2015.

[AHF+98] John Ashburner, Chloe Hutton, Richard Frackowiak, Ingrid Johnsrude, Cathy Price, and Karl Friston. Identifying global anatomical differences: Deformation-based morphometry. *Human Brain Mapping*, 6(5-6):348–357, 1998.

[AKLM03] Dmitri Alekseevsky, Andreas Kriegl, Mark Losik, and Peter W. Michor. The Riemannian geometry of orbit spaces. the metric, geodesics, and integrable systems. *Publ. Math. Debrecen*, 62, 2003.

[BC11] Jeremie Bigot and Benjamin Charlier. On the consistency of Fréchet means in deformable models for curve and image analysis. *Electronic Journal of Statistics*, (5):1054–1089, 2011.

[BG10] Jeremie Bigot and Sebastien Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Annals of Statistics*, 38(4):2422–2464, 2010.

[DATP17a] Loïc Devilliers, Stéphanie Allassonnière, Alain Trouvé, and Xavier Pennec. Inconsistency of template estimation by minimizing of the variance/pre-variance in the quotient space. *Entropy*, 19(6):288, 2017.

[DATP17b] Loïc Devilliers, Stéphanie Allassonnière, Alain Trouvé, and Xavier Pennec. Template estimation in computational anatomy: Fréchet means in top and quotient spaces are not consistent. *SIAM Journal on Imaging Sciences*, 10(3):1139–1169, August 2017.

[DBD+14] Hugo Darmanté, Benoit Bugnas, Regis Bernard De Dompsure, Laurent Barresi, Nina Miolane, Xavier Pennec, Fernand de Peretti, and Nicolas Bronsard. Analyse biométrique de l'anneau pelvien en 3 dimensions – à propos de 100 scanners. *Revue de Chirurgie Orthopédique et Traumatologique*, 100(7, Supplement):S241 –, 2014.

[DM98] Ian L. Dryden and Kanti V. Mardia. *Statistical shape analysis*. John Wiley & Sons, New York, 1998.

[Efr79] Bradley Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1):1–26, 1979.

[EJCB12] Alan C. Evans, Andrew L. Janke, D. Louis Collins, and Sylvain Baillet. Brain templates and atlases. *NeuroImage*, 62(2):911–922, August 2012.

[EM91] Michel Émery and Gabriel Mokobodzki. Sur le barycentre d'une probabilité dans une variété. *Séminaire de probabilités de Strasbourg*, 25:220–233, 1991.

[EPS12] Jürgen Ehlers, Felix A. E. Pirani, and Alfred Schild. Republication of: The geometry of free fall and light propagation. *General Relativity and Gravitation*, 44(6):1587–1609, 2012.

[GD04] John C. Gower and Garmt B. Dijksterhuis. *Procrustes problems*, volume 30 of *Oxford Statistical Science Series*. Oxford University Press, Oxford, UK, January 2004.

[GJR+14] David Gunther, Alec Jacobson, Jan Reininghaus, Hans-Peter Seidel, Olga Sorkine-Hornung, and Tino Weinkauf. Fast and Memory-Efficienty Topological Denoising of 2d and 3d Scalar Fields. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2585–2594, December 2014.

[GMT00] Alexandre Guimond, Jean Meunier, and Jean-Philippe Thirion. Average Brain Models: A Convergence Study. *Computer Vision and Image Understanding*, 77(2):192–210, February 2000.

[Goo91] Colin Goodall. Procrustes Methods in the Statistical Analysis of Shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339, 1991.

[Gow75] John C. Gower. Generalized procrustes analysis. *Psychometrika*, 40:33–51, 1975.

[HHLAP16] Mehdi Hadj-Hamou, Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: A Computational Framework Based on Stationary Velocity Fields. *Frontiers in Neuroscience*, June 2016.

[JDJG04] Sarang Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:151–160, 2004.

[JKSM06] Shantanu Joshi, David Kaziska, Anuj Srivastava, and Washington Mio. Riemannian structures on shape spaces: A framework for statistical inferences. In *Statistics and Analysis of Shapes*, Modeling and Simulation in Science, Engineering and Technology, pages 313–333. Birkhaeuser Boston, 2006.

[Joh66] Richard M. Johnson. The minimal transformation to orthonormality. *Psychometrika*, 31:61–66, 1966.

[JWS12] Alec Jacobson, Tino Weinkauf, and Olga Sorkine. Smooth shape-aware functions with controlled extrema. *Computer Graphics Forum (Proc. SGP)*, 31(5):1577–1586, July 2012.

[Ken77] David G Kendall. The diffusion of shape. *Advances in applied probability*, 9(3):428–430, 1977.

[Ken84] David G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[KM97] John T. Kent and Kanti V. Mardia. Consistency of procrustes estimators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):281–290, 1997.

[KSW11] Sebastian A. Kurtek, Anuj Srivastava, and Wei Wu. Signal estimation under random time-warpings and nonlinear signal alignment. In J. Shawe-taylor, R.s. Zemel, P. Bartlett, F.c.n. Pereira, and K.q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 675–683, 2011.

[LAP15] Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. Regional flux analysis for discovering and quantifying anatomical changes: An application to the brain morphometry in Alzheimer's disease. *NeuroImage*, 115:224–234, July 2015.

[Le98] Huiling Le. On the consistency of procrustean mean shapes. *Advances in Applied Probability*, 30(1):53–63, March 1998.

[Lel93] Subhash Lele. Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Mathematical Geology*, 25:573–602, 1993.

[LK93] Huiling Le and David G. Kendall. The Riemannian structure of euclidean shape spaces: A novel environment for statistics. *Annals of Statistics*, 21(3):1225–1271, 09 1993.

[MCAG93] Michael I. Miller, Gary E. Christensen, Yali Amit, and Ulf Grenander. Mathematical textbook of deformable neuroanatomies. *Proceedings of the National Academy of Sciences*, 90(24):11944–11948, December 1993.

[MHP17] Nina Miolane, Susan Holmes, and Xavier Pennec. Template Shape Estimation: Correcting an Asymptotic Bias. *SIAM Journal on Imaging Sciences*, 10(2):808 – 844, 2017.

[MHP18] Nina Miolane, Susan Holmes, and Xavier Pennec. Topologically Constrained Template Estimation via Morse–Smale Complexes Controls Its Statistical Consistency. *SIAM Journal on Applied Algebra and Geometry*, 2(2):348–375, January 2018.

[Mos39] Charles I. Mosier. Determining a simple structure when loadings for certain tests are known. *Psychometrika*, pages 149–162, 1939.

[MP13] Lorenzi Marco and Xavier Pennec. Parallel transport with pole ladder: Application to deformations of time series of images. In *Geometric Sci-*

ence of Information: First International Conference, GSI 2013, Paris, France, August 28-30, 2013. Proceedings, pages 68–75. Springer Berlin Heidelberg, 2013.

[MV96] J. B. Antoine Maintz and Max A. Viergever. An overview of medical image registration methods. Technical report, In Symposium of the Belgian hospital physicists association (SBPH-BVZF, 1996.

[MWP+07] Daniel S. Marcus, Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open access series of imaging studies (OASIS): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience.*, 19:1498–1507, 2007.

[Pen18] Xavier Pennec. Parallel Transport with Pole Ladder: a Third Order Scheme in Affine Connection Spaces which is Exact in Affine Symmetric Spaces. *ArXiv e-prints 1805.11436*, May 2018.

[Sch66] Peter H. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31:1–10, 1966.

[Sma96] Christopher G. Small. *The Statistical Theory of Shape*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[TB77] Jos M. F. Ten Berge. Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, 42:267–276, 1977.

[Zie77] Herbert Ziezold. *On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces*, pages 591–602. Springer Netherlands, Dordrecht, 1977.